IMPACT OF CORRELATING CER RISK DISTRIBUTIONS On A "REALISTIC" COST MODEL

By Alfred Smith, Shu-Ping Hu
Tecolote Research Inc.

ABSTRACT

Risk and risk correlation papers generally use simplified cost models to demonstrate the message of the paper. This paper uses a more typical cost model in terms of complexity, number of elements and cost methodologies. The model was developed in Automated Cost Estimating Integrated Tools (ACEIT) to illustrate a process for developing a cost estimating relationship (CER), selecting an appropriate risk distribution, defining its bounds and assigning an appropriate level of correlation between a CER and other CERs within the cost estimate. An Excel utility is used to extract Latin Hypercube simulation information when running the risk analysis and then use it to construct the correlation matrix that is achieved in the risk simulation for selected elements of the work breakdown structure (WBS). Using this utility, we are able to establish the correlation that exists in the model due to the functional relationships built into the model and before the analyst applies correlation. The study then demonstrated that the correlation developed in the simulation matches the correlation input by the analyst. A systematic comparison is made to illustrate how correlation between the WBS cost elements and the cost risk results are influenced as the analyst applies: correlation between WBS cost elements, introduces risk distributions to the CER inputs, introduces correlation amongst the CER input risk distributions and finally, adds a low level of correlation across all remaining elements (some programs require this assumption). The results illustrated in the paper serve to help the cost analyst visualize the impacts of his/her decisions as risk assumptions are layered into a model.

The paper also provides an overview of how the user defines correlation within the ACE model and insight into the ACE Group Strength algorithm (the method used within ACE to achieve the user defined correlation). However, the focus of the paper is on the results and their interpretation.

INTRODUCTION

The purpose of this paper is to examine the impact of correlation on a cost estimate's risk results. Applying correlation can be a real challenge. What does the analyst do, for instance, if correlation is known at one level but the model is built several levels lower? Completing an entire correlation matrix for a ten element WBS may be feasible…but what if there are 500 or a 1000 elements? What about the functional relationships in the model? Should the analyst's applied correlation take precedence? This paper will address some, but not all of these issues by way of a demonstration.

To set the stage for this discussion the key steps in creating the estimate will be reviewed first, namely: building the work breakdown structure (WBS); populating the WBS with cost estimating relationships (CERs); assigning risk distributions and bounds to each CER and their inputs including any penalty factors for schedule and technical complexity; and finally assigning appropriate correlations amongst the CERs and their inputs.

WORK BREAKDOWN STRUCTURE

Figure 1 illustrates the first four levels of a cost model created in the Automated Cost Estimator (www.aceit.com). This model is used as the basis for the study presented in this paper.

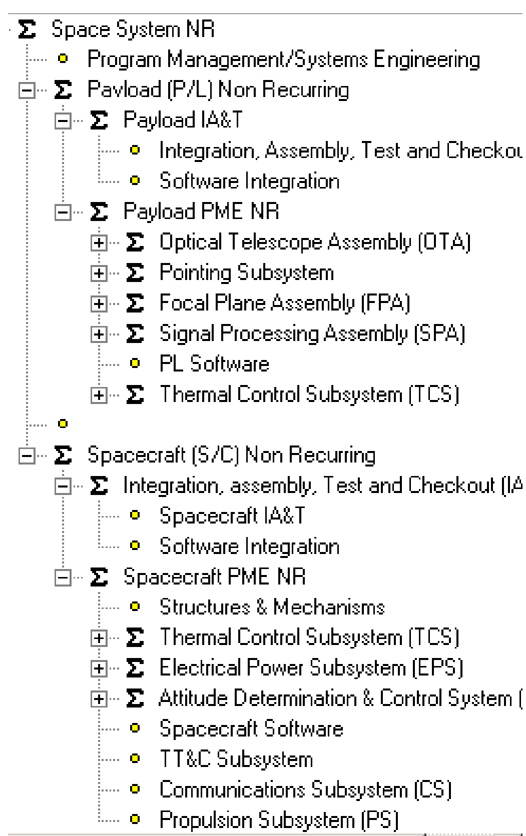


Figure 1: Sample Cost Model WBS

COST ESTIMATING RELATIONSHIPS

Populating the lowest levels of the WBS elements with the appropriate cost estimating relationships (CERs) is the next step. For this estimate, the WBS is populated with CERs from the Unmanned Space Vehicle Cost Model (USCM). For further information on the USCM CERs, you can request access at https://www.uscm8.com/Default.asp.

Figure 2 shows the data available at the USCM web site. In addition to the information shown, the "Documents" folder contains an USCM model created in ACE that one can download to run within the ACE Executive. If you do not have ACEIT, there is also a version of the model that you can run via your browser (ACE not required on the local machine).

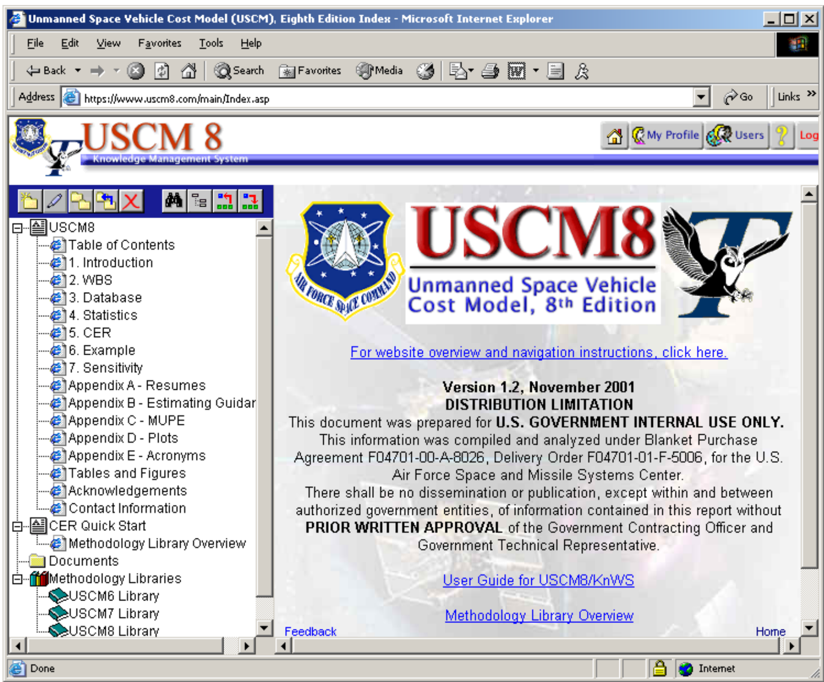


Figure 2: USCM Data Available on the Web

If you are unable to gain access to the **Figure 2** web site, there is a public domain web site http://www.jsc.nasa.gov/bu2/PCEHHTML/pceh220.htm that provides a dated but still relevant overview of how the USCM CERs have been developed.

The regression analysis process that yields the CER will also yield an objective measure of the CER error for a specified confidence level and position within the source data set. **Table 1** contains typical results that are available when a CER is generated in CO$TAT (the statistics application in the ACEIT suite of tools). The bounds on the estimate for a specific confidence level are influenced by it's the distance from the data set center of gravity. Note that the standard error of the CER is greater at the lower and upper bounds than mid range of the data.

Table 1: Typical Regression Analysis Results from CO$TAT

Unit$ = 92.93 + 27.39 * Wgt R^2Adj = 96% Mean Absolute Deviation = 7%	**Lower Bound of Data**	**Mid Range of Data**	**Upper Bound of Data**
Input Weight (lbs)	4	15	23
Confidence Level (%)	80%	80%	80%
Point Estimate Standard Error	**48.5**	**44.6**	**48.6**
Lower Bound	$134	$441	$654
Estimate	$202	$504	$723
Upper Bound	$271	$567	$792
	Bounds For RI$K		
Lower Bound	66%	87%	90%
Upper Bound	134%	113%	110%

Figure 3 illustrates the 80% and 99% confidence interval bounds on a CER. Note that one of the data points is outside the 80% confidence level. The analyst can use CO$TAT to calculate uncertainty bounds based on inputs associated with the alternative under investigation.

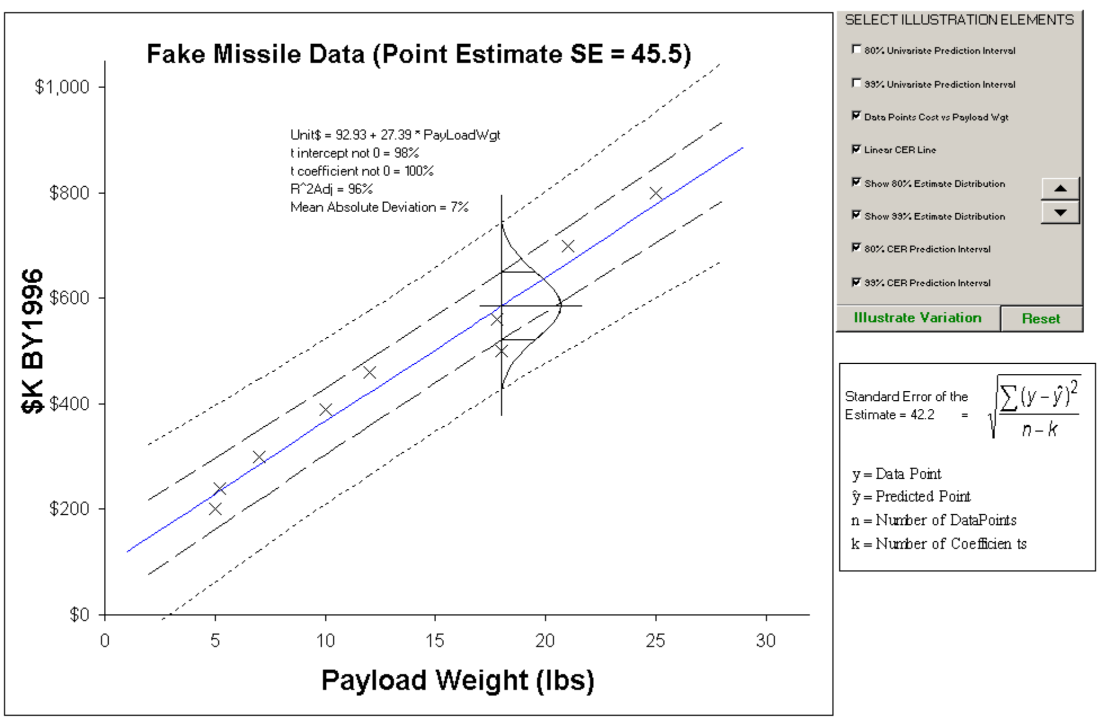


Figure 3: Accuracy of the Cost Estimating Relationship

CHOOSING RISK DISTRIBUTIONS AND BOUNDS

Having developed a CER, the analyst must now choose a distribution shape and ensure it is properly defined in order to model the uncertainty associated with the CER. If a CER is developed in CO$TAT, the user will be able to paste it and its documentation into the ACE model. During this process, ACE will apply a normal distribution to model risk for CERs containing an additive error term (e.g. linear) and log-normal for CERs containing a multiplicative error term (e.g. non-linear). The analyst is then free to make a subjective selection from any of the other available distribution forms: uniform, triangular, beta, or weibull.

The most common method used in ACE to define the distribution shape is to set the upper and lower bounds. We recommend that users enter the bounds as a % of the point estimate. That way all the bounds will move consistent with the point estimate under consideration. ACE requires the user to also specify the confidence level that the bounds represent. ACE will derive the entire distribution from this information. **Figure 4** illustrates how the shape and spread of a triangular distribution is modeled in ACE based upon bounds that are 70% of the point estimate for the low and 160% for the high and their interpretation is either 0-100%, 10-90% or 5-95%. Note, the interpretation of low and high bounds need not be symmetrical.

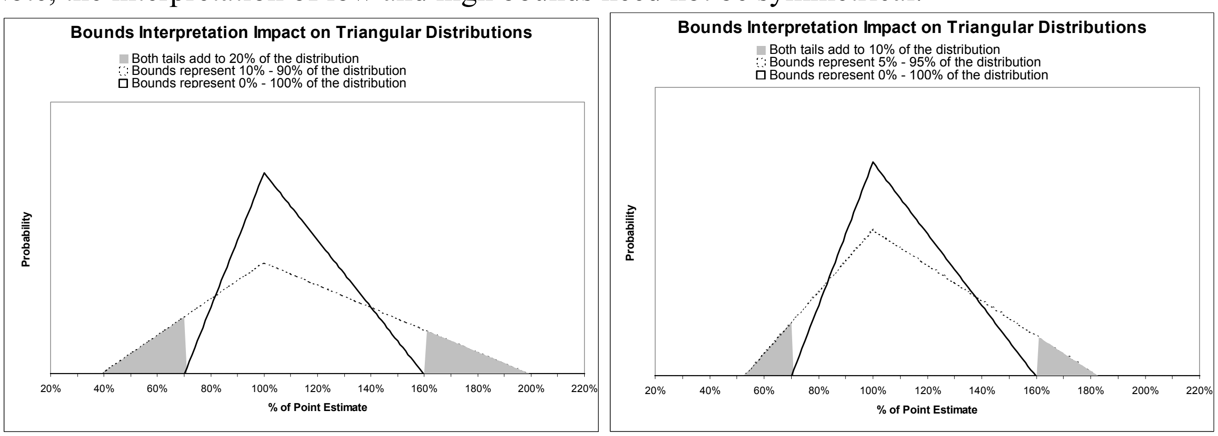


Figure 4: Impact of Bounds Interpretation on How a is Distribution Modeled

CO$TAT will generate statistically derived prediction intervals for use as the bounds of the distribution based upon user defined CER input(s) and a selected confidence level. To calculate the bounds information for export to ACE, CO$TAT uses the following equation:

$$Adj.\,SE = SE * \sqrt{1 + \frac{1}{n} + \frac{\left(\frac{x_0 - \bar{x}}{S_x}\right)^2}{n}} = SE * \sqrt{1 + \frac{1}{n} + \frac{\left(\frac{\text{distance}}{sample\ std}\right)^2}{n}}$$

where:

x_0 = the value of the independent variable used in the estimate
$\bar{x}$ = the mean of the independent variable in database
S_x = uncorrected sample standard dev of the independent variable
n = the number of data points

The distance assessment need only be characterized in terms of a number of standard deviations the estimate cost driver value is from the center of the CER source driver data. For example, if you assess the distance to be within approximately 2 sample standard deviations of the driver variable, then the ratio (of "distance" to "sample std") becomes 2. However, the ACE user is not required to run the prediction interval calculation in CO$TAT. For simplicity, we also establish some default values to address the assessment of this ratio based upon the similarities between the system to be estimated and the systems used to generate the CER:

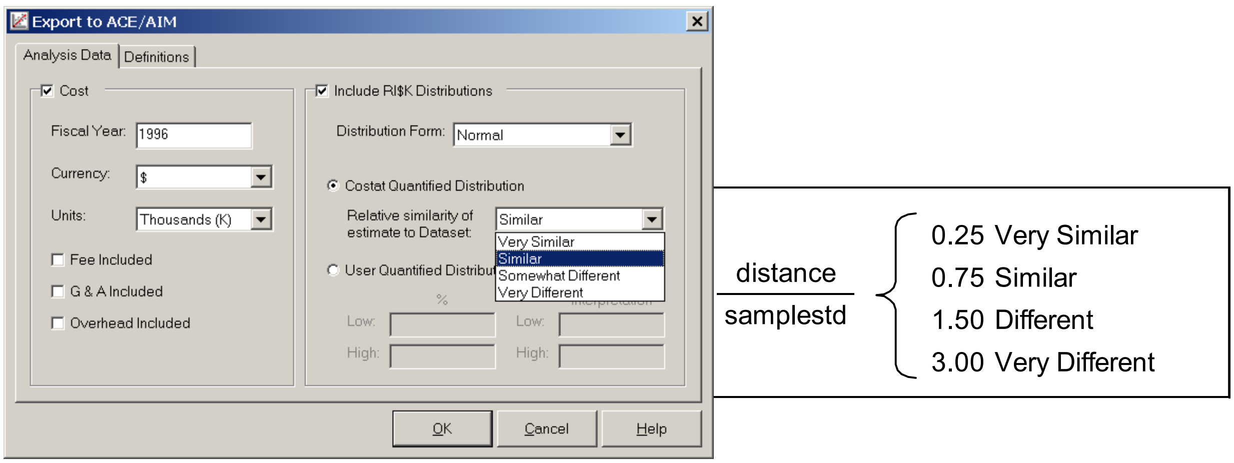


Figure 5: Establishing Risk in the ACE Estimate

Figure 5 illustrates the CO$TAT user interface to set the distribution shape and bounds. The dialog will open with defaults based on the CER type but allow the analyst to make a subjective assessment on shape, bounds and interpretation. For greater control, the user can enter specific bounds/interpretation based upon a prediction interval calculation conducted in CO$TAT. **Figure 6** shows RI$K options and some cost element settings for one of the study models.

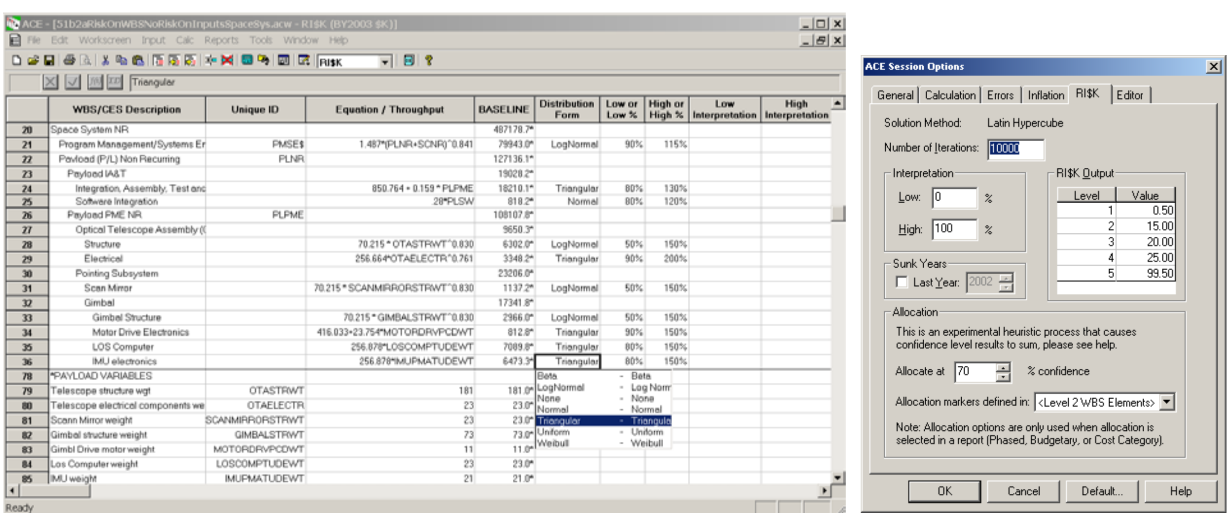


Figure 6: Establishing Risk in the ACE Estimate

CHOOSING PENALTY FACTORS

Schedule and technology risk assumptions can be modeled as penalty factors in RI$K. These factors affect the distribution tail on the high end. Penalty factors can be used to simulate the effect on program cost estimates of extraordinary schedule and technical difficulties. This is germane primarily to development efforts, except that there are a few occasions in production efforts where these considerations are important.

The problem in deriving penalty factors that can reflect independent and separate assessments of schedule and technical difficulties is that there are no normative, quantitative measures of schedule or technical difficulty, either separately or in combination. It is difficult to separate cost/schedule overrun into causative sources of technical and pure schedule difficulties as in the performance of a project technical difficulties manifest themselves to a great degree in schedule extensions that have cost consequences. The multipliers provided as guidance in ACE Help are recommended for use in cases where the analyst has no better information upon which to base the choice of a multiplier to capture the effects of extraordinary schedule or technical uncertainty. Our studies strongly indicate that each major class of hardware should have its own set of multipliers, but that would require a major research effort to develop highly quantitative factors. If an analyst has relevant information in a specific risk assessment, we recommend the use of that information to specify the penalty factors rather than reliance on the defaults.

There is another matter involving the use of these factors in the RI$K model: the requirement for penalty factors to be applied element-by-element throughout the WBS. Penalty factors are more correctly applied at the bottom line. If they are applied selectively to elements in the WBS so that some elements are not modified by these factors, then the effect at the bottom line will be less than if the factor had been applied at the bottom line. This suggests that the factors for individual elements should even be greater in magnitude, or that once the level of uncertainty has been decided upon, the factors should be applied uniformly to all WBS elements. The analyst must choose the application method that best reflects his knowledge of the case. In our case, we elected to not use any penalty factors.

CORRELATED RISK DISTRIBUTIONS: THEORETICAL BASIS FOR THE ACE METHOD

To set the stage for the main focus of this paper, we will begin by reviewing the two commonly used correlation coefficients to measure the association between two random variables: Pearson's Product Moment Correlation and Spearman's Rank Order Correlation. We will then review the Laurie-Goldberg[1] method for simulating random numbers in a risk analysis and then describe the process used in ACE.

<u>Pearson's Product Moment Correlation</u>

The most common measure of correlation is Pearson's Product Moment Correlation (called Pearson's correlation for short). Pearson's correlation reflects the degree of linear relationship between two sets of numbers. The formula for Pearson's correlation takes on many forms. A few commonly used formulas are shown below:

$$r = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{\sqrt{\sum (X-\bar{X})^2}\sqrt{\sum (Y-\bar{Y})^2}} \quad \textbf{or} \quad r = \frac{n\sum XY - \sum X \sum Y}{\sqrt{\left(n\sum X^2 - \left(\sum X\right)^2\right) * \left(n\sum Y^2 - \left(\sum Y\right)^2\right)}}$$

or

$$r = \frac{\sum Z_X Z_Y}{n} \qquad \text{where:} \qquad Z_X = \frac{X-\mu}{\sigma}$$

n = number of ordered pairs
σ = standard deviation
μ = mean
X = first variable of an ordered pair
Y = second variable of an ordered pair

ACEIT uses the Pearson's definition to model correlations in risk simulations.

Spearman's Rank Order Correlation

Spearman's rank order correlation is used as a nonparametric test statistic to determine if two random variables are independently distributed. However, the correlation formula is applied to the ranks rather than the original variable values. Therefore, no assumptions are made about the underlying distributions. When Spearman's correlation is significantly different from zero, it can be interpreted that there is an association between two variables, just like the ordinary Pearson product-moment correlation.

To calculate the Spearman Rank Order correlation, replace the value of each X by the value of its rank compared to all the other X's in the sample. The resulting list of numbers will be members of a perfectly known distribution, namely the uniform distribution of integers between 1 and N, inclusive. If some of the ranks are identical, all of the "ties" are assigned the mean of the ranks that they would have had if their values had been slightly different. In this situation, some of the ranks could be partial integers. In all cases the sum of all assigned ranks will be the same as the sum of the integers from 1 to N, namely N(N + 1)/2. Let R_i be the rank of X_i among the other X's, S_i be the rank of Y_i among the other Y's, ties being assigned the appropriate midrank as described above. The rank-order correlation coefficient is defined to be the linear correlation coefficient of the ranks, and is computed as follows:

$$r_s = \frac{\sum_i \left[\left(R_i - \bar{R}\right) * \left(S_i - \bar{S}\right)\right]}{\sqrt{\sum_i \left(R_i - \bar{R}\right)^2} * \sqrt{\sum_i \left(S_i - \bar{S}\right)^2}}$$

If there are no ties in the ranking, then the equation can be reduced to:

$$r_s = 1 - \frac{6 * \sum d^2}{n * (n^2 - 1)}$$

where:

d = difference between the paired ranks
n = number of paired ranks

Both Crystal Ball and @Risk use *Spearman's rank order* correlation to model correlations that is useful if the underlying distributions are not normal. Rank order correlation is also easier to simulate than Pearson's correlation, but it is not appropriate for cost risk analyses[2]. Spearman's correlation measures the monotonicity between two random variables. For example, if two random variables are perfectly increasing monotonically as $Y = X^{10}$, then the Spearman's correlation between X and Y is 1.0. This measure is not suitable for dealing with the correlations that may exist in the work breakdown structures (WBS) in our business because the variance of the sum (or any linear transformation) of the random variables used to create the cost estimate is related to the Pearson's correlation, not the Spearman's correlation.

Lurie-Goldberg's Simulation Method[1]

The Lurie-Goldberg algorithm uses *Pearson's product moment* correlation in the simulation process and starts from a method known as Cholesky's decomposition. Let us assume a set of n variables, Y_1, Y_2, … and Y_n, are associated with the marginal distribution functions, F_1, F_2, … and F_n respectively, and they are also correlated by the following correlation matrix:

$$Cov(\mathbf{Y}) = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & . & \rho_{1n} \\ \rho_{12} & 1 & \rho_{23} & . & \rho_{2n} \\ \rho_{13} & \rho_{23} & 1 & . & \rho_{3n} \\ . & . & . & . & . \\ \rho_{1n} & \rho_{2n} & \rho_{3n} & . & 1 \end{pmatrix}$$

The main steps in the Lurie-Goldberg's (LG) method are given below:

1. Generate n independent draws, Z1, Z2, … and Zn, from a standard normal distribution, i.e., N(0,1).
2. Construct n correlated standard normal random variables X_1, X_2, … and X_n using Cholesky's factorization matrix (L):

$$X_1 = Z_1$$
$$X_2 = \rho_{12}\, Z_1 + \sqrt{1-\rho_{12}{}^2}\; Z_2$$
$$X_3 = \rho_{13}\, Z_1 + \frac{\rho_{23} - \rho_{12}\rho_{13}}{\sqrt{1-\rho_{12}{}^2}}\, Z_2 + \frac{\sqrt{(1-\rho_{12}{}^2)(1-\rho_{13}{}^2) - (\rho_{23} - \rho_{12}\rho_{13})^2}}{\sqrt{1-\rho_{12}{}^2}}\, Z_3$$
$$\ldots$$

(i.e., X = LZ such that LL' = Cov(Y))

3. Apply the standard normal CDF function to each X variable to derive the corresponding uniform correlated draws:

$$U_i = \int_{-\infty}^{X_i} \frac{1}{\sqrt{2\pi}}\, e^{-\frac{y^2}{2}}\, dy$$

4. Invert the above uniform draws by the user-defined marginal distribution Fi:

$$Y_i = F_i^{-1}(U_i)$$

5. Iterate over the Cholesky's factorization matrix (L), go back to step 2, and repeat the process until the desired level of correlations is achieved.

In step 2, correlation is introduced by using Cholesky's decomposition method to a set of independent normal random variables. After applying Cholesky's method, these correlated normal random variables are then transformed into correlated uniform random variables using their cumulative distribution functions (see step 3 above). The correlated uniforms are then converted into the correlated random variables of the desired marginal distribution functions. Since non-linear transformations are involved in the process, the correlations among the WBS elements will differ from the correlation among the normal distributions constructed by Cholesky's method. The final step is to iterate on the Cholesky's factors to achieve the desired correlations among the WBS elements. This way, both the marginal distributions (i.e., the shape of the distribution: triangular, beta, log-normal, etc.) and the correlation structure are preserved.

ACE/RI$K Simulation Method

ACE uses a modified Lurie-Goldberg algorithm to create a set of random variables that match the user-supplied correlations. The differences between ACE and the Lurie-Goldberg methods are given below:

1. ACE only allows the user to enter a single vector of correlation coefficients where the correlations are relative to the dominant cost driver in a particular "Group" of WBS elements. By doing this, the remaining members of the correlation matrix are "implied" (and therefore consistent) and the algorithm is simplified.

2. ACE uses ranks during the simulation process to smooth out the resulting variables to make them suitable for the Latin-Hypercube (LH) simulation. Ranking in this context is for the purpose of generating the LH draws such that they closely resemble the original input distributions, and it should **not** be confused with rank order correlation.

3. ACE does not iterate on the user supplied "Group Strengths" to achieve the desired correlations among the WBS elements. Nonetheless, in our test cases the user-defined group strengths match the desired correlations very closely, all within 0.5%.

The general steps for the ACE/RI$K algorithm are listed below for the purpose of comparing with the LG method. We will use the general assumptions as given above except Y_1 is now assumed to be a dominant item in the group and the pairwise correlation between Y_i and Y_1 is ρ_i for i = 2, …, n.

1. Generate n independent draws, Z_1, Z_2, … and Z_n, from a standard normal distribution, i.e., N(0,1).

2. Construct n correlated standard normal random variables X_1, X_2, … and X_n using Cholesky's pairwise factorization formula:

$$
\begin{aligned}
&X_1 = Z_1 \\
&X_2 = \rho_2\, Z_1 + \sqrt{1-\rho_2{}^2}\, Z_2 \\
&X_3 = \rho_3\, Z_1 + \sqrt{1-\rho_3{}^2}\, Z_3 \\
&\ldots \\
&X_n = \rho_n\, Z_1 + \sqrt{1-\rho_n{}^2}\, Z_n
\end{aligned}
$$

   (**Note:** If no dominant item is identified within a given group and Y_1 is assigned by a correlation of ρ_1, then $X_1 = \rho_1\, Z_1 + \sqrt{1-\rho_1{}^2}\, Z_1'$, just as the rest of the X variables. Here Z_1' is a random draw for X_1 and Z_1 is a random draw for this group)

3. Generate the corresponding uniform LH draws for the X_i variables consistent with the normal cumulative probability for each of the X_i values.

4. Invert the above uniform draws by the user-defined marginal distribution Fi:

$$Y_i = F_i^{-1}(U_i)$$

ACE requires the user to specify only a vector of pairwise correlations across the WBS elements of a specific group. In this way, a consistent matrix is generated that simplifies the LG method (removes requirement to test for consistency). ACE starts like the LG method by taking the random draws and applying the Cholesky's method. However, instead of using the normal CDF to derive the percentiles associated with the correlated normal draws, ACE orders the normal draws to achieve the uniformed correlated draws (see step 3 above). The final step of inverting the uniform draws by the user-defined marginal distribution is the same.

This method is simple, there is no need to iterate, marginal distributions are preserved, and the desired correlations are achieved. Our tests do not reveal significant differences in results if you go to the trouble of defining a complete, even inconsistent, correlation matrix.

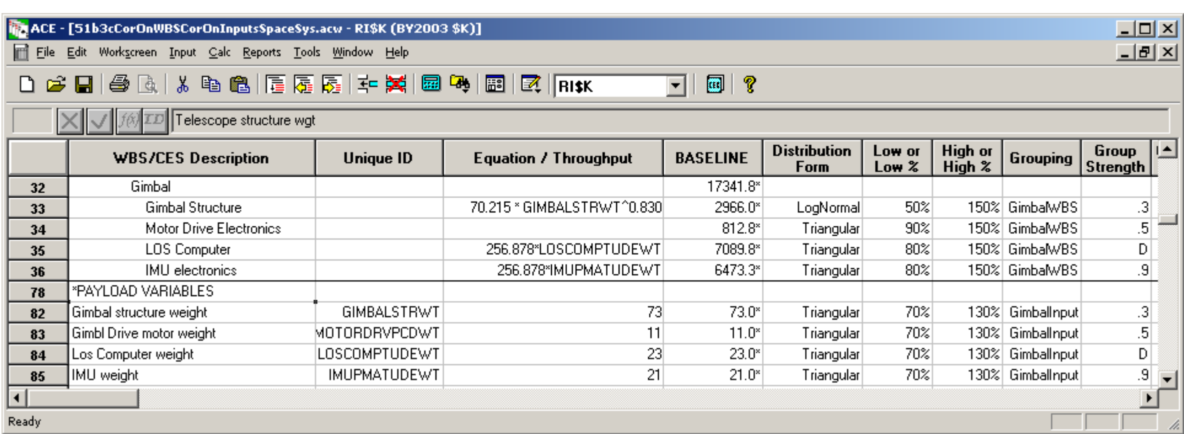
CORRELATED RISK DISTRIBUTIONS: A STUDY CONDUCTED IN ACE

A small section in the cost model was singled out for detailed analysis. **Table 2** identifies the case names used in this study and the risk assumptions associated with them.

Table 2: Systematic Layering of Risk Assumptions

	Cost Method Risk			Configuration Risk		
Case Name	**Risk on All CERs**	**Correlation on Selected CER Risk**	**Low Correlation on all other CERs**	**Risk on All Inputs**	**Correlation on Selected Input Risk**	**Low Correlation on all other Inputs**
NoRiskOnWBSRiskOnInputs				X		
NoRiskOnWBSCorOnInputs				X	X	
NoRiskOnWBSCorOnInputs+Low				X	X	X
RiskOnWBSNoRiskOnInputs	X					
RiskOnWBSRiskOnInputs	X			X		
RiskOnWBSCorOnInputs	X			X	X	
CorOnWBSNoRiskOnInputs	X	X				
CorOnWBSRiskOnInputs	X	X		X		
CorOnWBSCorOnInputs	X	X		X	X	
CorOnWBSCorOnInputs+LowWBS	X	X	X	X	X	
CorOnWBSCorOnInputs+LowAllElse	X	X	X	X	X	X

Figure 6 illustrates how the "CorOnWBSCorOnInputs" case was modeled. A "D" is assigned to the dominant element in the Group ("GimbalWBS" for the CERs, "GimbalInputs" for the weight inputs). The remaining values in the Group Strength column are the ***pairwise correlations*** between an element and the associated dominate element in the Group.



	WBS/CES Description	Unique ID	Equation / Throughput	BASELINE	Distribution Form	Low or Low %	High or High %	Grouping	Group Strength
32	Gimbal			17341.8*					
33	Gimbal Structure		70.215 * GIMBALSTRWT^0.830	2966.0*	LogNormal	50%	150%	GimbalWBS	.3
34	Motor Drive Electronics			812.8*	Triangular	90%	150%	GimbalWBS	.5
35	LOS Computer		256.878*LOSCOMPTUDEWT	7089.8*	Triangular	80%	150%	GimbalWBS	D
36	IMU electronics		256.878*IMUPMATUDEWT	6473.3*	Triangular	80%	150%	GimbalWBS	.9
78	*PAYLOAD VARIABLES								
82	Gimbal structure weight	GIMBALSTRWT	73	73.0*	Triangular	70%	130%	GimbalInput	.3
83	Gimbl Drive motor weight	MOTORDRVPCDWT	11	11.0*	Triangular	70%	130%	GimbalInput	.5
84	Los Computer weight	LOSCOMPTUDEWT	23	23.0*	Triangular	70%	130%	GimbalInput	D
85	IMU weight	IMUPMATUDEWT	21	21.0*	Triangular	70%	130%	GimbalInput	.9

Figure 7: Assigning Correlation to Selected Elements

The ACE methodology assumes that the remaining elements in the full correlation matrix will be the product of the associated inputs (Gimbal Structure to IMU electronics would be 30% * 90% = 27%). This is a simplifying assumption to be sure, but in our tests, a reasonable one.

Correlations Input Into ACE

Gimbal				
Gimbal Structure	100%	15%	30%	27%
Motor Drive Electronics		100%	50%	45%
LOS Computer			D	90%
IMU electronics			90%	100%

Figure 8: Complete Correlation Matrix Developed by ACE.

The gray cells **Figure 8** represent the correlation entered into the model. To measure correlation achieved in the ACE risk simulation, Tecolote created a utility in Excel (available upon request) to capture user selected results for each iteration of the risk simulation. The Excel function "CORREL" is used to calculate the Pearson Correlation coefficient between selected elements in the estimate and build the associated full correlation matrix. **Figure 9** shows the 3 screens in this utility populated with the results from one of the cases defined in **Table 2.**

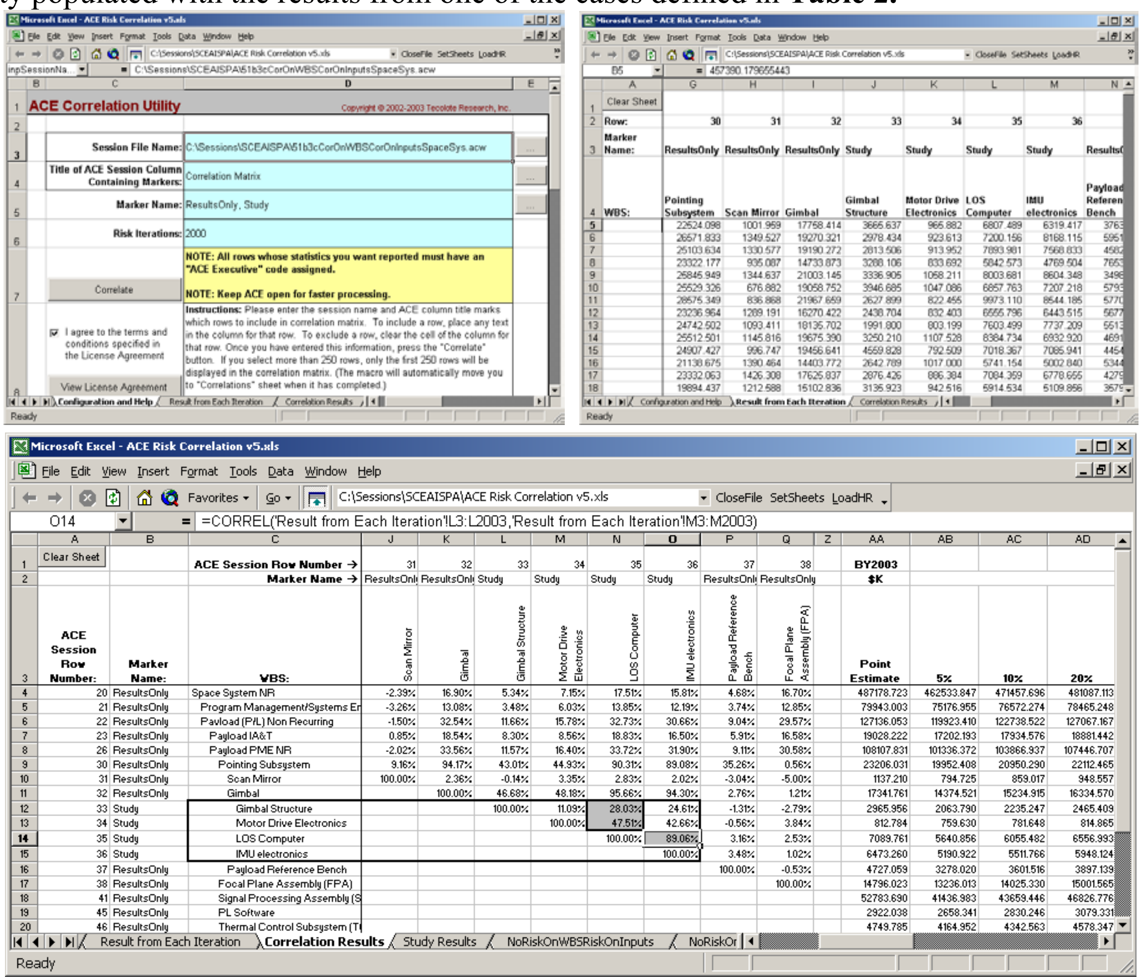


Figure 9: ACE Correlation Utility

The ACE methodology differs from others in that ACE does not "iterate" through the simulation. The final step of the ACE method may "change" the correlation values, i.e. the correlations between the X_i variables are "correct," but those between the Y_i variables are different if the relationship between X_i and Y_i is not linear. This may be a reason why methods such as Laurie-Goldberg iterate to a solution. The ACE Correlation Utility was used to determine if this was a problem in the way ACE implements correlation and if not, how many risk simulation iterations are required to approach the input correlations.

Correlations Input Into ACE

Gimbal				
Gimbal Structure	100%	15%	30%	27%
Motor Drive Electronics		100%	50%	45%
LOS Computer			D	90%
IMU electronics			90%	100%

25 Iterations

Gimbal				
Gimbal Structure	100%	15%	-8%	-8%
Motor Drive Electronics		100%	43%	48%
LOS Computer			100%	88%
IMU electronics				100%

100 Iterations

Gimbal				
Gimbal Structure	100%	8%	10%	8%
Motor Drive Electronics		100%	33%	30%
LOS Computer			100%	89%
IMU electronics				100%

500 Iterations

Gimbal				
Gimbal Structure	100%	13%	27%	23%
Motor Drive Electronics		100%	46%	42%
LOS Computer			100%	91%
IMU electronics				100%

2000 Iterations

Gimbal				
Gimbal Structure	100%	11%	28%	25%
Motor Drive Electronics		100%	48%	43%
LOS Computer			100%	89%
IMU electronics				100%

10000 Iterations

Gimbal				
Gimbal Structure	100%	15%	29%	26%
Motor Drive Electronics		100%	46%	41%
LOS Computer			100%	90%
IMU electronics				100%

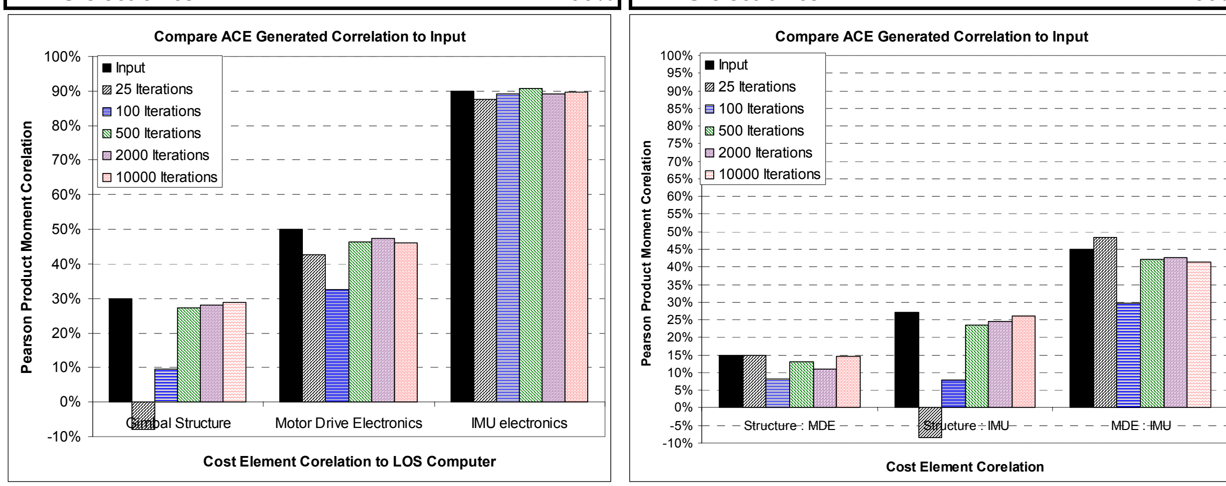


Figure 10: Tabular and Graphical Comparison of Correlation Input to Output

Figure 10 illustrates how the developed correlation compares to the inputs as the number of simulation iterations is increased. 500 iterations appear sufficient to replicate the input vector, however 2000 or more may be required in order to establish the implied correlations. The conclusion we have drawn from this and many other tests is that ACE does in fact replicate the intended correlations reasonably well in most cases. The results for the Motor Drive (46% at 10k iterations rather than the input 50% gives us pause and may be another indicator why iterating may be required if greater accuracy is needed.

Figure 11, shows how the cumulative cost probability curve for various iterations compares to the 10k iteration cost risk result. Risk is applied on 35 cost rows (four are correlated under the Gimbal element, the remaining 30 are assigned a low correlation) and 31 input weights (four driving the Gimbal children are correlated, the remaining 27 are assigned a low correlation). Note that for the Gimbal (specific correlation applied) and for the total cost (mix of correlations involved) there is little difference between the 500, 2000 and 10k iteration cost results.

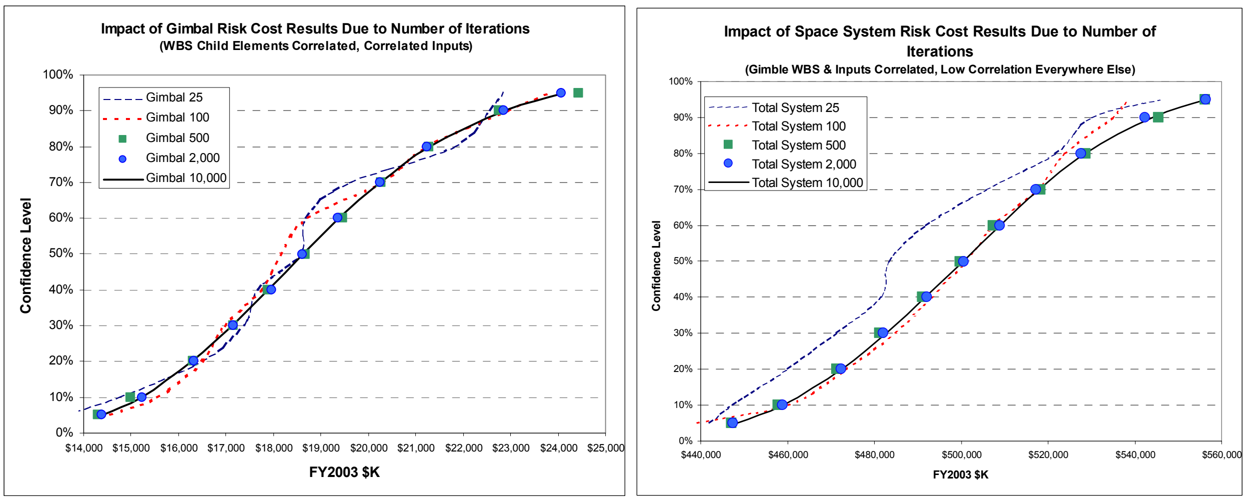


Figure 11: Cost Results Using Different Number of Simulation Iterations

The ACE Correlation Utility results show no functional correlations in the Gimbal cost elements. As seen in **Figure 12**, the correlation matrix is populated with values less than a few percent for the case where risk is applied to both CERs and inputs.

RiskOnWBSRiskOnInputs				
Gimbal Structure	100%	0%	0%	0%
Motor Drive Electronics		100%	0%	0%
LOS Computer			100%	-1%
IMU electronics				100%

Figure 12: Correlation Results in the Model Before Group Strength is Applied

Figure 13 illustrates how correlation that is applied at the input level only and then at the CER level only is manifested at the CER level in the risk simulation

RiskOnWBSCorOnInputs				
Gimbal Structure	100%	3%	10%	9%
Motor Drive Electronics		100%	13%	11%
LOS Computer			100%	40%
IMU electronics				100%

CorOnWBSNoRiskOnInputs				
Gimbal Structure	100%	15%	30%	28%
Motor Drive Electronics		100%	49%	44%
LOS Computer			100%	90%
IMU electronics				100%

Figure 13: Correlation Results in the Model After Group Strength is Applied

Figure 14 illustrates the correlation developed as risk assumptions are layered on the estimate. For instance, if no risk is applied at the CER (WBS) level, the correlation applied at the input level is in fact developed at the CER level. Once risk at the CER level is added, the correlation at the CER level is no longer the same as that applied at the inputs. **This suggests that small CER correlations measured in data sets could in fact be due to the correlation of inputs rather then the CERs.**

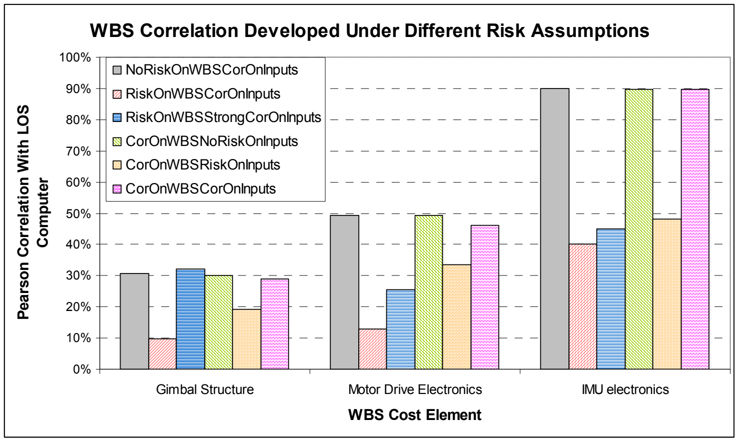


Figure 14: Correlation Results Under Different Risk Assumptions

Figure 15 illustrates how the risk cost results for the Gimbal changes as the risk assumptions are layered on the estimate (10k iterations). The risk estimate is symmetrical about the 50% confidence level when the only risk applied is at the input level. The chart also demonstrates that for this estimate, cost risk is far more significant than configuration risk.

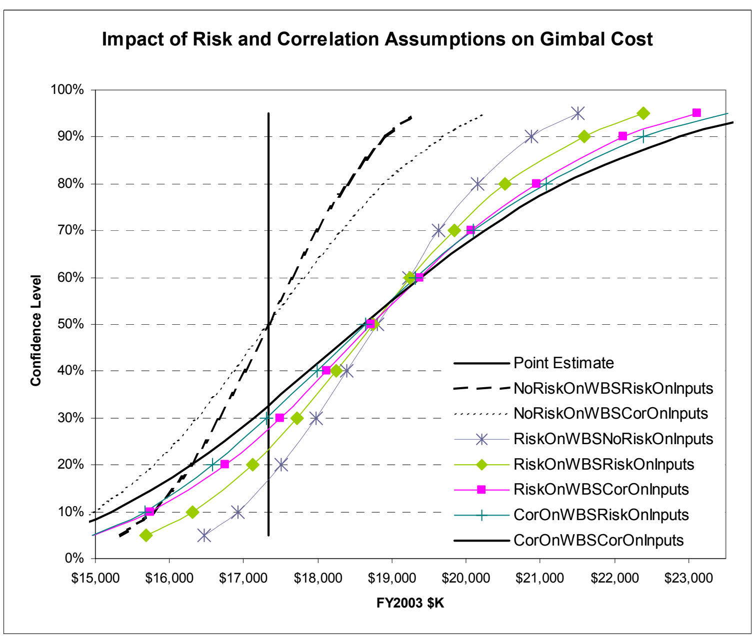


Figure 15: Gimbal Cost Results Under Different Risk Assumptions

Figure 16 illustrates the results at the total cost level under different risk assumptions (10k iterations). In this model, risk on the CERs has a greater impact than the risk on the inputs. Interestingly, the point estimate "improves" from 12% to 30% confidence level when risk is applied to the inputs. But the range in possible costs broadens significantly. Low level correlation across all elements other than those associated with the Gimbal does have an impact, but the impact at the 80% confidence level is only a few percent.

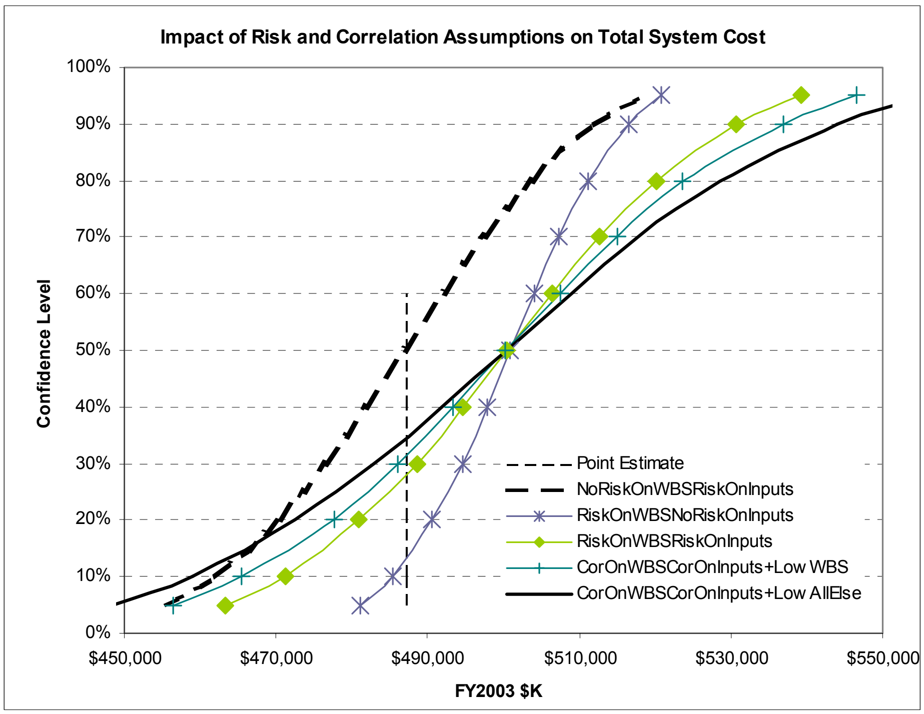


Figure 16: Total Cost Results Under Different Risk Assumptions

CONCLUDING REMARKS

Correlations applied at the WBS level will be affected if risk is then applied at the input level. If correlations are found in the historical cost data, the analyst should not overlook that possibility that the cost correlations are in fact the result of cost driver correlations.

The ACE Correlation Utility will give the analyst confidence that the correlations achieved in the ACE RI$K simulation are reasonably close to those that are input. Tecolote has completed dozens of simple model comparisons with tools such as Crystal Ball and @Risk using a variety of distribution and correlation assumptions (including full, inconsistent matrices in other tools). In every case the ACE cost results match the other tools within a few percent at the 10% or 90% confidence level. For the cost model studied in this paper, it was evident that correlation did not have much of an impact at the total cost level. Models with hundreds of rows and/or greater correlation may see a greater impact.

BIOGRAPHIES

Alfred Smith: General Manager for the Software Products/Services Group. Over 20 years experience leading, executing or contributing to life cycle cost model development for wide variety systems such as Aircraft Carriers, Submarines, Multi Mission Aircraft, Amphibious Assault Vehicles, International Space Station and Customs Modernization. 10 years experience teaching cost estimating, statistical analysis and risk modeling. Tecolote lead for the development, distribution and support of the ACEIT suite of tools. Education: 1986 University College, London, England, Master of Science with Distinction; 1978 Royal Military College, Kingston, Bachelor of Engineering.

Address:
Alfred B. Smith
Tecolote Research, Inc.
5266 Hollister Ave. No 301
Santa Barbara, CA 93111

Tel: (805) 964-6963
Fax: (805) 964-7329
E-mail: asmith@tecolote.com

Dr. Shu-Ping Hu: Principal analyst at Tecolote Research, Inc. Tecolote's expert in all statistical matters since 1984 and principal resource for the design, development, modification, and integration of statistical software packages Tecolote has developed to conduct regression analysis, learning curve derivation/implantation, and cost risk analysis. Over 15 years experienced developing independent cost estimates, cost analysis research, and risk analysis implementation. Advocated an iterative regression technique (MUPE) to model a multiplicative error term without bias, developed correction factors (PING Factor) to adjust for the downward bias in log-error models and principal developer of the correlation method implemented in the ACE simulation method. Education: 1975 National Taiwan University, BS – Mathematics; 1981 University of California at Santa Barbara, MS – Statistics; 1983 University of California at Santa Barbara, Ph. D. – Statistics

Address:
Shu-Ping Hu
Tecolote Research, Inc.
5266 Hollister Ave. No 301
Santa Barbara, CA 93111

Tel: (805) 964-6963
Fax: (805) 964-7329
E-mail: shu@tecolote.com

REFERENCES

1. Simulating Correlated Random Variables; Philip M. Lurie and Matthew S. Goldberg; Institute for Defense Analyses; 32nd DODCAS; 2-5 February 1999

2. Do Not Use RANK CORRELATION in Cost Risk Analysis; Paul R. Garvey; MITRE Corporation, 32nd DODCAS; 2-5 February 1999

3. Why Correlation Matters in Cost Estimating; Stephen A. Book; The Aerospace Corporation; 32nd DODCAS; 2-5 February 1999